

# Collaborative Research Practices and Shared Infrastructures for Humanities Computing

Maristella Agosti  
Francesca Tomasi  
(Eds)



00001010110  
00010101101010  
0001010110  
0010101101010  
0101011010  
0010101101010  
000101011  
0001010110101

**Collaborative Research Practices  
and Shared Infrastructures  
for Humanities Computing**

2<sup>nd</sup> AIUCD Annual Conference, AIUCD 2013  
Padua, Italy, 11-12 December 2013

Proceedings of Revised Papers

Maristella Agosti and Francesca Tomasi (Eds)

cleup

ASSOCIAZIONE per  
l'INFORMATICA UMANISTICA  
e la CULTURA DIGITALE



<http://www.umanisticadigitale.it>

Prima edizione: settembre 2014

ISBN 978 88 6787 260 2

CLEUP sc  
"Coop. Libreria Editrice Università di Padova"  
via G. Belzoni 118/3 – Padova (t. 049 8753496)  
[www.cleup.it](http://www.cleup.it) - [www.facebook.com/cleup](https://www.facebook.com/cleup)

© 2014 AIUCD

Tutti i diritti di traduzione, riproduzione e adattamento,  
totale o parziale, con qualsiasi mezzo (comprese  
le copie fotostatiche e i microfilm) sono riservati.

In copertina:  
Graphic Design: Massimo Malaguti – Scuola Italiana Design (elaborazione del logo  
di AIUCD)

## Organization

AIUCD 2013 was organized by the Information Management Group of the  
Department of Information Engineering of the University of Padua, Italy.

### Committees

#### General Chair

*Dino Buzzetti*, Presidente AIUCD

#### Program Chairs

*Maristella Agosti*, Dipartimento di Ingegneria dell'Informazione, Università  
degli Studi di Padova

*Anna Maria Tammaro*, Dipartimento di Ingegneria dell'Informazione, Uni-  
versità degli Studi di Parma

#### Program Committee

*Fabio Ciotti*, Dipartimento Studi Umanistici, Università di Roma Tor Vergata  
*Giorgio Maria Di Nunzio*, Dipartimento di Ingegneria dell'Informazione,  
Università degli Studi di Padova

*Maurizio Lana*, Dipartimento di Studi Umanistici, Università del Piemonte  
Orientale

*Federico Meschini*, Dipartimento di Scienze Umanistiche, della Comunica-  
zione e del Turismo, Università degli Studi della Tuscia

*Nicola Orio*, Dipartimento di Beni Culturali, Università degli Studi di Padova  
*Nicola Palazzolo*, già ordinario nell'Università di Perugia

ASSOCIAZIONE per  
l'INFORMATICA UMANISTICA  
e la CULTURA DIGITALE

<http://www.umanisticadigitale.it>

Prima edizione: settembre 2014

ISBN 978 88 6787 260 2

CLEUP sc  
"Coop. Libreria Editrice Università di Padova"  
via G. Belzoni 118/3 – Padova (t. 049 8753496)  
[www.cleup.it](http://www.cleup.it) - [www.facebook.com/cleup](http://www.facebook.com/cleup)

© 2014 AIUCD

Tutti i diritti di traduzione, riproduzione e adattamento,  
totale o parziale, con qualsiasi mezzo (comprese  
le copie fotostatiche e i microfilm) sono riservati.

In copertina:  
Graphic Design: Massimo Malaguti – Scuola Italiana Design (elaborazione del logo  
di AIUCD).

## Organization

AIUCD 2013 was organized by the Information Management Group of the Department of Information Engineering of the University of Padua, Italy.

### Committees

#### General Chair

*Dino Buzzetti*, Presidente AIUCD

#### Program Chairs

*Maristella Agosti*, Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova

*Anna Maria Tammaro*, Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Parma

#### Program Committee

*Fabio Ciotti*, Dipartimento Studi Umanistici, Università di Roma Tor Vergata  
*Giorgio Maria Di Nunzio*, Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova

*Maurizio Lana*, Dipartimento di Studi Umanistici, Università del Piemonte Orientale

*Federico Meschini*, Dipartimento di Scienze Umanistiche, della Comunicazione e del Turismo, Università degli Studi della Tuscia

*Nicola Orio*, Dipartimento di Beni Culturali, Università degli Studi di Padova

*Nicola Palazzolo*, già ordinario nell'Università di Perugia

*Roberto Rosselli Del Turco*, Dipartimento di Studi Umanistici, Università di Torino

*Marco Rufino*, Fondazione Rinascimento Digitale, Firenze

*Francesca Tomasi*, Dipartimento di Filologia Classica e Italianistica, Università di Bologna

#### **Award Chair**

*Francesca Tomasi*, Dipartimento di Filologia Classica e Italianistica, Università di Bologna

#### **Local Committee**

*Debora Leoncini*, Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova

*Marta Manfioletti*, Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova

*Chiara Ponchia*, Dipartimento di Beni Culturali, Università degli Studi di Padova

*Gianmaria Silvello*, Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova

## **Table of contents**

PREFACE / PREFERENZA	11
<i>Maristella Agosti, Francesca Tomasi</i>	

#### **KEYNOTE**

KEYNOTE ADDRESS / INTERVENTO INVITATO

Toward a Computational Narratology	17
<i>Jan Christoph Meister</i>	

#### **INVITED**

CONTRIBUTIONS FROM RESEARCH GROUPS AND CENTERS /

CONTRIBUTI DI CENTRI E GRUPPI DI RICERCA

Nuovi scenari per la ricerca in filosofia: i testi e gli strumenti del portale Daphnet	39
<i>Michela Tardella, Cristina Marras</i>	

Acquisizione e Creazione di Risorse Plurilingui per gli Studi di Filologia Classica in Ambienti Collaborativi	55
<i>Federico Boschetti</i>	

Da <i>Musisque Deoque</i> a <i>Memorata Poetis</i> . Le vie della ricerca intertestuale	69
<i>Paolo Mastandrea, Luigi Tessarolo</i>	

**PANELS**DIGITAL RESOURCES AND NETWORK SERVICES FOR DIGITAL HUMANITIES  
RESEARCH / RISORSE DIGITALI E SERVIZI DI RETE PER LA RICERCA

## IN CAMPO UMANISTICO

*Digital humanities*: difficoltà istituzionali e risposte infrastrutturali 81  
*Dino Buzzetti*

*Digital humanities* e analisi dei testi 89  
*Paolo Mastandrea*

Infrastrutture e risorse digitali. L'esperienza dell'ILIESI 93  
*Antonio Lamarra*

DH@ILC: linee di attività e ricerca 101  
*Simonetta Montemagni*

THE DIGITAL LIBRARY TO SUPPORT THE COMPUTER HUMANIST /  
LA BIBLIOTECA DIGITALE A SUPPORTO DELL'UMANISTA INFORMATICO

Digital libraries and digital humanities scholars: community context,  
workflow and collaboration 115  
*Anna Maria Tammaro*

e-Infrastructures per le esigenze della ricerca 121  
*Rossella Caffo*

(Formal) Models for systems, infrastructures, communities,  
and cultures 129  
*Nicola Ferro*

Biblioteche digitali e studi umanistici 135  
*Maurizio Lana*

Some remarks about Museo Galileo's digital collections 143  
*Stefano Casati, Fabrizio Butini*

**PAPERS**

## DIGITAL PHILOLOGY / FILOGIA DIGITALE

L'*Open Philology Project* dell'Università di Lipsia. Per una filologia  
'sostenibile' in un mondo globale 151  
*Monica Berti, Greta Franzini, Emily Franzini,*  
*Giuseppe Celano, Gregory R. Crane*

A collaborative tool for philological research: experiments  
on Ferdinand de Saussure's manuscripts 163  
*Angelo Mario Del Grosso, Simone Marchi, Francesca Murano,*  
*Luca Pesini*

Edition Visualization Technology: a tool to publish digital editions 177  
*Raffaele Masotti, Julia Kenny*

Codifying the codex. The digital edition of the *Becerro Galicano*  
of San Millán 187  
*David Peterson*

## DIGITAL CULTURAL HERITAGE / PATRIMONIO CULTURALE DIGITALE

ASIt: Atlante Sintattico d'Italia: A linked open data geolinguistic  
web application 197  
*Giorgio Maria Di Nunzio, Jacopo Garzonio, Diego Pescarini*

The "Verbo-Visual Virtual" Platform for Digitizing and Navigating  
Cultural Heritage Collections 205  
*Alessandro Marchetti, Sara Tonelli, Roberto Sprugnoli*

*Dante*. A Web Application for the History of Art 219  
*Chiara Ponchia*

Digital Lightbox: a web-based visualization framework applied  
to paleographical research 229  
*Giancarlo Buomprisco*

Towards a shared methodology for audio preservation:  
Luciano Berio's private collection of sound recordings 237  
*Federica Bressan, Sergio Canazza*

Knowledge objects and bodies of knowledge: knowledge sharing  
platforms applied to international relations 249  
*Giuseppe Vitiello*

## EDUCATIONAL APPROACHES / DIDATTICA

Moodle as a collaborative platform for digital humanities 261  
*Giuseppe Fiorentino, Maria Accarino,*  
*Alessia Pierfederici, Daniela Rotelli*

Geostoria del quotidiano. Proposte per un'analisi automatica del testo letterario <i>Alessia Scacchi</i>	269
Managing Educational Information on University Websites: a proposal for Unibo.it <i>Federico Nanni</i>	279
Author index	287

## Preface

The Italian Association for Digital Humanities<sup>1</sup> (AIUCD) was launched in 2011 to promote and disseminate methods and facilitate scientific collaboration and development of useful resources in the field of digital humanities in Italy. AIUCD is an associate organization of the European Association for Digital Humanities<sup>2</sup> (EADH), which brings together and represents the Digital Humanities in Europe across the entire spectrum of disciplines that research, develop, and apply digital humanities methods and technology. AIUCD is thereby represented in the Alliance of Digital Humanities Organizations<sup>3</sup> (ADHO) which promotes and supports digital research and teaching across all arts and humanities disciplines, acting as a community-based advisory force, and supporting world-wide excellence in research, publication, collaboration and training.

We are very pleased to present the volume of the proceedings of the 2<sup>nd</sup> Annual Conference of the Italian Association for Digital Humanities (AIUCD 2013) on “Collaborative Research Practices and Shared Infrastructures for Humanities Computing”<sup>4</sup>, which took place at the Department of Information Engineering of the University of Padua, 11-12 December 2013. The first conference was held in Florence, 13-14 December 2012, for “An Agenda for Humanities Computing and Digital Culture”<sup>5</sup>. The third conference

<sup>1</sup> Associazione per l'Informatica Umanistica e la Cultura Digitale <http://www.umanisticadigitale.it/>

<sup>2</sup> <http://www.eadh.org/>

<sup>3</sup> <http://adho.org/>

<sup>4</sup> <http://aiucd2013.dei.unipd.it/>

<sup>5</sup> <http://www.umanisticadigitale.it/2012/12/>

## Da *Musisque Deoque* a *Memorata Poetis* Le vie della ricerca intertestuale\*

Paolo Mastandrea, Luigi Tessarolo

Dipartimento di Studi Umanistici, Università Ca' Foscari, Venezia, Italia  
mast@unive.it, luigi.tessarolo@fastwebnet.it

**Abstract.** Le concordanze, gli onomastici, i volumi di indici tematici e le raccolte di luoghi simili a stampa sono ormai da qualche decennio sostituiti dall'interrogazione automatica dei grandi *corpora* digitalizzati. L'applicazione delle tecnologie informatiche in campo filologico non può limitarsi ad ampliare gli archivi, ma deve aprire nuove vie per rispondere a domande sempre più esigenti sulla storia, la circolazione e il riuso dei testi, antichi e moderni. Si dà qui una panoramica delle funzionalità di ricerca metrico-verbale in *Musisque Deoque* (<http://www.mqdq.it>) e delle componenti in sviluppo nella nuova risorsa dedicata alla ricerca tematica-semantica *Memorata poetis* (<http://www.memoratapoetis.it>).

**Parole chiave:** intertestualità formale e concettuale, memoria poetica, archivi di testi letterari.

### 1. Introduzione

Da quando – ormai una quindicina di anni orsono – iniziammo a caricare testi di poesia latina in rete, il fuoco della nostra attenzione era rivolto al raffinamento della ‘ricerca intertestuale’ – un affare che ha preso questo nome grazie a Julia Kristeva (1969)<sup>1</sup>, ma può dirsi costituitosi oggetto di pratica quotidiana sin dai primordi della disciplina filologica, nata negli ambienti della Biblioteca d’Alessandria, durante il regno di Tolomeo II Filadelfo e dunque verso la metà del III secolo a.C.

\* M. Agosti, F. Tomasi (Eds). *Collaborative Research Practices and Shared Infrastructures for Humanities Computing*. 2nd Aiucd Annual Conference, Aiucd 2013. CLEUP, Padova, 2014.

<sup>1</sup> La letteratura sulla metodologia e la pratica della intertestualità sta diventando smisurata, e quasi incontrollabile, giorno dopo giorno; rimando per questo alla recente sintesi di Neil Coffee, un latinista statunitense che si muove allineato sulla direzione e verso gli scopi che cerchiamo di perseguire coi nostri studi.



Il nostro interesse al problema corre su due piani complementari. Al primo livello ci si pone in linea di continuità con le vecchie concordanze a stampa, che agevolavano le ricerche dei singoli vocaboli seguendo elenchi alfabetici iniziali; ovviamente, sarebbe superfluo spiegare quale vantaggi ulteriori aggiunga l'interrogazione automatica nel reperimento di legami che coinvolgono gli abbinamenti, o le associazioni di più parole, in sequenza o a distanza, nell'ordine stesso o in quello inverso, oppure limitatamente a parti di esse: rimane pur sempre principio fondamentale dell'esplorazione l'esame di una stringa di caratteri alfabetici.

Al secondo livello si prosegue piuttosto l'antica pratica dell'indicizzazione tematica, a caccia di quella che Cesare Segre chiamava 'interdiscorsività'<sup>2</sup> – utile ad uno studio dei sintagmi diegetici: trame e intrecci del racconto, per lo più ma non esclusivamente in prosa. In un caso e nell'altro, si tratta di una continuazione della obsoleta *Quellenforschung*; la critica di impianto positivistico vi faceva corrispondere *grosso modo* due diverse tipologie di studi, che ad esempio rispetto alle fonti del capolavoro ariostesco produsse con Pio Rajna (*Le fonti dell'Orlando Furioso*, Firenze 1876<sup>1</sup>) un capostipite dell'indagine condotta sulle genealogie tematiche, mentre Antonio Romizi (*Le fonti latine dell'Orlando Furioso*, Torino 1896) si concentrava sugli apparati dei riferimenti puntuali alla tradizione letteraria, classica e italiana; o ancora, contrapporrei i saggi ricostruttivi di Eduard Norden (il grande commentatore di *Vergils Aeneis VI*, Leipzig 1903<sup>1</sup>, ma soprattutto restauratore di parti degli *Annales* col suo *Ennius und Vergilius*, Leipzig 1915), a fronte delle tante raccolte di *loci similes*, così accuratamente redatte in quei decenni non solo da giovani aspiranti filologi nelle loro dissertazioni, ma anche da accademici maturi e affermati in tutti i campi – e vorrei citare il repertorio di *Scripture and Classical Authors in Dante*, curato da Edward Moore e uscito per la prima volta ad Oxford nel 1896. C'è da restare stupiti di fronte all'ampiezza della documentazione, in rapporto ai mezzi allora disponibili: eppure la reazione di matrice idealistica ebbe allora la capacità di stroncare un intero filone di studi<sup>3</sup>.

Ma torniamo all'oggi, magari guardando al domani. Attualmente si procede con pari interesse – se non uguale distribuzione di forze – su due linee

<sup>2</sup> Segre 1982.

<sup>3</sup> Per un inquadramento storico e bibliografico del dibattito avviato in Italia dal saggio del Croce su *La ricerca delle fonti* (in *Problemi di Estetica*, Bari, Laterza, 1910) si veda Pasini 1988, pp. 7-30.

direttrici: non parallele, anzi abbastanza divergenti fra loro, ma promettenti entrambe di buoni risultati

## 2. La ricerca metrico-verbale estesa alle varianti dei testi latini in *Musisque Deoque*<sup>4</sup>

Alle caratteristiche iniziali – e più volte illustrate, anche in sede di convegni e relativi atti a stampa<sup>5</sup> – altre non secondarie funzioni si sono aggiunte negli ultimi mesi. Alludiamo innanzi tutto alla integrazione con <http://www.pedecerto.eu>, programma di metrica latina digitale, le cui risorse sono ancora da sfruttare sia al fine di condurre ricerche automatiche di forme / lemmi con analogia collocazione nel verso, sia per disambiguare su base prosodica nei testi le occorrenze di forme che possono appartenere a più lemmi.

Introdotta più di recente, ancora sotto collaudo, è la funzione <http://www.mqdq.it/cooccorrenze.jsp>, che riesce a confrontare un testo sorgente, a scelta del ricercatore, con l'intero corpus di *Musisque Deoque*, oppure con una sua parte; al momento le restrizioni del target consentite sono le seguenti: 1) solo autori precedenti; 2) solo autori successivi; 3) solo metri dattilici (default); però sarà presto possibile limitare il target ad uno o più autori / opere.

Il confronto consiste nella ricerca delle cooccorrenze di coppie di parole del testo sorgente nei testi target, con le seguenti opzioni selezionabili dall'utente:

- distanza tra le due parole, misurata come numero di parole interposte: da 0 a 5 (default 1);
- ricercare le parole solo nella stessa sequenza o anche in ordine inverso (default);
- escludere o meno termini molto frequenti (sono contenuti in una lista di 121 voci, al momento non modificabile, ma lo si potrebbe fare);
- ricercare o meno (default) anche le varianti in apparato;
- ricercare solo le forme esatte oppure cercare per lemma (default).

<sup>4</sup> Allocati negli archivi e interrogabili attraverso motore di ricerca al sito: <http://www.mqdq.it>

<sup>5</sup> Da ultimo, in breve: *Text Retrieval on Critical Editions*, a cura di M. Manca, Linda Spinazzè, F. Boschetti oltre a chi scrive, in *Workshop on Annotation of Corpora for Research in the Humanities* (Heidelberg, Jan. 5, 2012), pp. 1-12 = «Journal for Language Technology and Computational Linguistics» 26, 2011, pp. 129-140.

Per quanto riguarda lo score, l'assegnazione di un punteggio a ciascuna occorrenza non ha lo scopo di individuare i migliori risultati, in maniera da farli affiorare e presentarli allo studioso per primi (un po' come fa Google), quanto piuttosto di eliminare i peggiori. Il fatto è che anche un breve componimento produce, se messo a confronto con l'intero corpus poetico (diverso sarebbe il confronto tra due singoli autori o opere), un numero molto elevato di risultati (mediamente assai più di 100 per singolo verso, con i parametri di default) e i criteri adottati per giudicarli non hanno una grana abbastanza fine da riuscire a portarne in superficie una quantità sufficientemente ristretta da costituire una guida per lo studioso. In altri termini, se il testo sorgente non è proprio di dimensioni minime, le occorrenze a punteggio massimo sono comunque molte centinaia o migliaia, e il nostro algoritmo non ha la pretesa di effettuare tra esse una più fine discriminazione. Perciò, in attesa di realizzare uno strumento di maggiore capacità risolutiva (e intravediamo la possibilità di raggiungere un tale risultato ricorrendo a confronti metrici fini, mettendo a frutto, come si diceva, le opportunità di *Pedecerto*), abbiamo ritenuto più vantaggiosa per lo studioso una esposizione dei risultati di tipo tradizionale, cioè nell'ordine dei versi del testo sorgente, sfruttando comunque lo scoring per rendergli la vita più facile eliminando le occorrenze meno significative.

Lo strumento è pensato in modo speciale per i testi poetici; questo vale particolarmente per i criteri della ricerca per lemmi e per quelli dello scoring. Nei dettagli, la ricerca per lemmi non è estesa a tutte le forme flesse che fanno capo al lemma, ma solo a quelle con lo stesso numero di sillabe della forma sorgente. Per converso la query può estendersi pure ad altri lemmi, nel caso di voci composte che differiscano per il solo prefisso.

I criteri per l'assegnazione di un punteggio all'occorrenza sono i seguenti:

- 1) identità delle forme: 2 punti per parola;
- 2) stessa sequenza delle due parole: 2 punti;
- 3) stessa distanza tra le parole nel testo sorgente e nel target: 2 punti
- 4) posizione nel verso: è assegnato 1 punto per ciascuna parola che occupa, nei testi confrontati, la stessa posizione significativa (prima, seconda, penultima, ultima); sono dati complessivamente 2 punti anche nel caso in cui una sola parola condivida la posizione, ma sia uguale la distanza tra le parole.

Si trovano ancora in fase sperimentale, ma non proprio aurorale, le ricerche da effettuare mediante <http://www.mqdq.it/cometri.jsp>, predisposte alle funzioni:

- 1) Cerca lo schema del verso (tutte le pause, sinalefe, iato)
- 2) Cerca lo schema del verso (solo pause principali)
- 3) Cerca le parole nella posizione metrica
- 4) Cerca sequenze di 4/5 sillabe

Le prime due cercano i versi che riproducono per intero lo schema (quantità e pause) del verso sorgente, senza alcun riguardo alle parole o sillabe interessate.

Le funzioni 3 e 4 sono invece intese ad individuare riprese di sequenze metrico-verbali parziali, anche e soprattutto in assenza di una identità morfologica, basate perciò prevalentemente su una somiglianza di suono. Questi ultimi due approcci producono di fatto risultati parzialmente diversi e complementari, ma stiamo lavorando a combinare le due ricerche in un'unica operazione.

Intanto, con 3) si cercano le parole nella posizione metrica: da ciascun termine del verso sorgente è estratta una chiave di ricerca costituita dalla sua posizione metrica e dalle sole vocali, che viene poi reperita nel corpus; sono accolte solo le occorrenze che presentano la coincidenza di almeno quattro sillabe, in una o più parole; ma poi le occorrenze sono ordinate per rilevanza, sulla base del numero delle sillabe coincidenti, della contiguità delle parole trovate e della corrispondenza di consonanti.

Con 4) si cercano sequenze di quattro/cinque sillabe, in tal modo: si estraggono dal verso tutte le possibili sequenze di 4 o 5 sillabe, senza riguardo ai confini di parola; di ogni sillaba si considerano la posizione metrica e la vocale; dalla grande massa dei risultati si estraggono quelli che presentano un certo numero di corrispondenze anche nella parte consonantica delle sillabe.

### 3. La ricerca dei rapporti intertestuali di carattere semantico e tematico

Sin dalle origini della Filologia, cui si accennava, il transito e il reimpiego di elementi da un testo all'altro fu sempre visto come uno dei principali obiettivi di questa disciplina. Scopo fondamentale del progetto *Memorata Poetis*, che coinvolge studiosi di varie sedi universitarie e di altre istituzioni di ricerca nazionali (lo presentiamo qui, ma si vada anche direttamente al sito <http://www.memoratapoetis.it>) è la riunione, l'adattamento, il perfezionamento e per certa parte la creazione ex novo, di archivi digitali cui applicare strumenti elettronici di indagine capaci di analizzare testi plurilingui

con il fine di scorgerne le mutue riprese e relazioni. Si potrebbe così individuare ogni presenza (cosciente o incosciente) della memoria di un poeta all'atto della riscrittura di un altro, sicché quanto di solito è soltanto postulato, oppure anche documentato ma in modi occasionali ed estemporanei, trovi conferme oggettive – anzi, se è lecito usare il termine, – scientifiche'.

Il lavoro di esegesi, più o meno sistematica e completa dei testi, andrà così fiancheggiato, sostenuto e in più di un caso presumibilmente anticipato da una analisi fine dei dati, ottenuta incrociando due procedimenti: per un verso, nel tentativo di creare all'interno di una stessa lingua un thesaurus che permetta di eccedere i limiti della pura ricerca lessicale, esplorando i *corpora* testuali lemmatizzati con procedure (semi-)automatiche, comunque adeguate alle caratteristiche morfologiche, stilistiche, insomma linguistiche dei diversi testi (fatta salva una preventiva scelta di genere che garantisca livelli certi di ripetitività se non di formularità); basate sulla osservazione empirica per cui termini allocati in contesti simili hanno molto probabilmente significati simili o comunque correlati; nonché di stabilire relazioni reciproche dei termini fra lingue diverse, sfruttando dizionari bilingui greco-inglese, latino-inglese e arabo-inglese e grazie ad algoritmi probabilistici applicati ai contesti.

Dall'altro verso, la classificazione dei singoli documenti sulla base di tematiche, situazioni, ruoli e personaggi che popolano la produzione letteraria caratterizzata dalla maggiore brevità e concisione: l'epigrammatica, l'ecloga, l'elegia, la lirica e le corrispettive eredità (ad esempio in volgare italiano il sonetto di epoca tardomedievale, umanistica e seriore).

Si sta effettuando un censimento il più completo possibile di iscrizioni poetiche, antiche, medievali, umanistiche, moderne, e un trattamento di marcatura in vista della loro connessione a banche dati confluite già entro archivi elettronici di poesia latina, consultabili in rete <http://www.mqdq.it> e <http://www.poetiditalia.it> o su CD-Rom (*PoetriaNova*, 2da edizione, Firenze, SISMEL, 2010). Vi confluiscono materiali di tradizione diretta (da supporti materiali duri: pietra, bronzo, muro, ecc.) e indiretta (da manoscritti antologici e miscellanee, dall'interno di opere letterarie, ecc.), con relativi apparati critici, contesti di immagini, corredi di bibliografia. In tal modo il motore di interrogazione abbraccia una tradizione culturale più che millenaria con lo sguardo rivolto alla sopravvivenza delle tecniche espressive e stilistiche, curioso verso i meccanismi con cui agirono dall'interno i processi di riuso e attualizzazione; mirando ad una analisi combinata e raffinata del dettato e del portato del messaggio, offrendo ad una pluralità di studiosi i mezzi in-

vestigativi più opportuni – mezzi che divengono insuperabili o insostituibili quando siano applicati alla casistica dei prelievi non-consci degli autori dal deposito della memoria individuale.

Senza attendere tempi troppo lunghi, e men che meno la conclusione di un lavoro così imponente, sarà messo a disposizione della comunità scientifica, direttamente fruibile su web, un *corpus* di testi, per lo più ma non soltanto epigrafici, articolato per lingua (Latino, Greco, Italiano, Arabo) ed esteso in un arco cronologico dai primi documenti scritti dell'antichità ellenica ai secoli moderni; interrogabile sulla base di criteri storici, filologici, linguistici, stilistici; ed ancora, archeologici, paleografici, iconografici; sarà dunque offerta agli utenti della rete, attraverso appositi strumenti, la possibilità di individuare le correlazioni più latenti tra testi epigrafici e testi letterari. La marcatura previa dei testi darà luogo alla redazione di un lessico della letteratura epigrammatica, organizzato per aree semantiche, in sottoinsiemi che mettano in evidenza le relazioni analogiche esistenti fra gli elementi delle aree stesse; verrà così quasi automaticamente tracciata una mappa della circolazione o della migrazione, cronologica e topografica, dei temi e dei motivi individuati.

La componente più nettamente sperimentale del progetto comprende la messa a punto di strumenti e procedure informatiche in grado di rilevare in modo automatico (o semi-automatico) la presenza di temi e motivi all'interno delle composizioni poetiche, l'esplorazione degli spazi semantici grazie all'applicazione di metodi statistici al fine di identificare relazioni semantiche che emergono dalla cooccorrenza di termini in contesti simili, il confronto delle relazioni reciproche dei termini fra lingue diverse. I risultati attesi vanno dalla possibilità di ottenere una rappresentazione grafica bidimensionale delle distanze tra documenti epigrafici e documenti letterari, da cui poter inferire la probabilità che gli uni siano derivati dagli altri, o viceversa ne siano stati d'ispirazione (nel caso del confronto plurilingue è lecito attendersi analogamente che i documenti epigrafici processati nella lingua veicolare si dispongano sul piano cartesiano in modo da mostrare le relazioni intertestuali interlinguistiche), alla realizzazione di un metamotore di ricerca che travalichi sia i limiti lessicali dell'interrogazione verbale, sia i confini tra le culture (latina, greca, italiana) e i generi (letterario, epigrafico), ed eventualmente, oltre che operare su chiavi scelte di volta in volta dall'utente, possa essere impiegato per una scansione su larga scala dei testi al fine di riconoscere 'similitudini' fra i componimenti dei *corpora* epigrafici da un lato e quelli letterari dall'altro.

Le tecnologie e i formalismi che sono attualmente rubricati sotto la etichetta di Web Semantico mirano ad associare alle risorse informative variamente disponibili in rete una descrizione formale del loro significato, sicché un programma possa elaborare tale informazione in modo significativo (cioè tenendo conto di che cosa essa significhi), dedurne conseguenze, e generare automaticamente nuove informazioni. Questo nuovo orizzonte apre prospettive interessanti anche alle ricerche di tipo letterario, soprattutto nell'ambito della "tematologia" comparata e nello studio sui fenomeni della intertestualità, come testimoniato dalla letteratura internazionale più recente nel campo delle *Digital Humanities*. Lo stesso si può dire delle tecnologie di *knowledge extraction* e *text mining* applicate a vasti corpora testuali di tipo letterario, che sono state sperimentate con risultati interessanti in almeno tre progetti di ricerca in area statunitense, tra cui possiamo segnalare il *Tesserae Project: Intertextual Analysis of Latin Poetry* <http://tesserae.caset.buffalo.edu> di Neil Coffee *et al.*, il *Monk Project* <http://monkproject.org> diretto da John Unsworth e i progetti diretti da Franco Moretti presso lo Stanford Literary Lab.

La lemmatizzazione delle lingue classiche e l'attribuzione della categoria grammaticale si possono eseguire manualmente, ottenendo un grado di precisione molto elevato, oppure con l'applicazione di algoritmi probabilistici. Fin dall'origine della linguistica computazionale, il primo sistema fu applicato da p. Roberto Busa al *corpus Thomisticum* e dal CIPL-LASLA di Liegi a testi di autori della grecità e della latinità classica. *Morpheus*, il sistema di analisi morfologica realizzato da Gregory Crane ed usato presso il *Perseus Project* <http://www.perseus.tufts.edu>, è stato esteso da qualche anno con un sistema in grado di mettere in ordine decrescente di probabilità le possibili analisi proposte. Quanto al latino, il sistema di analisi morfologica realizzato a Pisa presso l'ILC-CNR – che si può dire rappresenti lo stato attuale dell'arte nel suo ambito – è disponibile online all'indirizzo <http://www.ilc.cnr.it/lemlat/lemlat/index.html>. Per l'arabo, si sta lavorando all'interno dello stesso Istituto, anche ai fini specifici della nostra ricerca.

L'analisi metrica automatica delle diverse realizzazioni dell'esametro greco e di altri metri non lirici può essere condotta facilmente con varie metodologie, mentre per i versi dattilici latini è già pronto uno strumento efficace e raffinato come *Pede certo* – di cui abbiamo parlato poco sopra.

L'esplorazione degli spazi semantici si effettua applicando metodologie statistiche a corpora testuali. I thesauri vengono generalmente definiti come collezioni ordinate di termini legati da relazioni di tipo gerarchico, asso-

ciativo e sinonimico, che costituiscono il lessico specialistico di un dominio della conoscenza e vengono usati per l'indicizzazione e il reperimento dell'informazione all'interno di tale dominio. Quelli basati su un sistema di classificazione multidimensionale dimostrano maggior flessibilità rispetto a quelli costruiti in base ad una classificazione monodimensionale, poiché forniscono una molteplicità di punti di accesso all'informazione e quindi maggior possibilità di indagini sui dati lessicali. Anche questi aspetti teorici e le varie applicazioni sono oggetto di lavoro per le due unità di Roma Tor Vergata e di Pisa, ILC-CNR.

#### 4. Bibliografia

- Coffee N. (2012). *Intertextuality in Latin Poetry*. In D. Clayman, ed., *Oxford Bibliographies in Classics*, OUP.
- Pasini G.F. (1988). *Dossier sulla critica delle fonti (1896-1909)*, Pàtron.
- Segre C. (1982). *Intertestuale / interdiscorsivo. Appunti per una fenomenologia delle fonti*. In C. Di Girolamo, I. Paccagnella, a c. di, *La parola ritrovata. Fonti e analisi letteraria*, Sellerio, pp. 15-28.

L'Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD) promuove e diffonde in Italia la riflessione metodologica e teorica, la collaborazione scientifica e lo sviluppo di pratiche, risorse e strumenti condivisi nel campo delle "Digital Humanities" (DH). AIUCD è associata alla "European Association for Digital Humanities" (EADH), che riunisce gli studiosi di DH in Europa ed è rappresentata nella "Alliance of Digital Humanities Organizations" (ADHO), che promuove e sostiene la ricerca digitale a livello mondiale. Questo volume raccoglie i contributi del secondo convegno annuale di AIUCD dedicato a ricerca collaborativa e piattaforme condivise per l'informatica umanistica. Il volume contribuisce ad arricchire il *corpus* di pubblicazioni scientifiche che rappresentano diverse aree del settore attive in Italia.

Maristella Agosti è professore ordinario di Ingegneria informatica presso il Dipartimento di Ingegneria dell'Informazione, Università di Padova, e socio corrispondente dell'Accademia Galileiana di Scienze Lettere ed Arti. È autore di numerosi contributi nel settore dei sistemi di gestione dell'informazione.

Francesca Tomasi è ricercatore di Ingegneria informatica presso il Dipartimento di Filologia Classica e Italianistica, Università di Bologna. È autore di numerose opere e saggi nel settore delle Digital Humanities.

ISBN 978 88 6787 260 2

